**Unit 6**

1. **Submit:** A link to a Google Doc (1-2 paragraphs) on your experience at the meetup. Discuss your experience with your career coach at your next call. DONE (for now)
2. Update LinkedIn Profile

DONE (for now)

**Unit 7**

**Project: Capstone Project 1: Data Story**

https://www.springboard.com/assets/images/icon-clock@2x.pngStudents typically spend 10 - 20 Hours

How do you create a data story? You’ve learned the basics, but the information is probably a bit abstract at this point. Keep in mind that storytelling is an art, so you have to get your imagination bubbling. In this project, you’ll learn some pointers to get those creative juices flowing. In the following sections, we’ll work step-by-step to create your first data story.   
  
**Steps:**

1. Ask the following questions and look for the answers using code and plots:
   1. Can you count something interesting?
   2. Can you find trends (e.g. high, low, increasing, decreasing, anomalies)?
   3. Can you make a bar plot or a histogram?
   4. Can you compare two related quantities?
   5. Can you make a scatterplot?
   6. Can you make a time-series plot?
2. Looking at the plots, what are some insights you can make? Do you see any correlations? Is there a hypothesis you’d like to investigate further? What other questions do the insights lead you to ask?
3. Now that you’ve asked questions, hopefully you’ve found some interesting insights. Is there a narrative or a way of presenting the insights using text and plots that tells a compelling story? What are some other trends/relationships you think will make the story more complete?

**Submission:** Submit links to a GitHub repository containing a Jupyter Notebook. The Notebook should contain:

* The questions you asked
* The trends you investigated
* The resulting visualizations and conclusions

You will be evaluated using this [rubric.](https://docs.google.com/document/d/1SPlg6_G7bnnbuEzpIK3Tc9pPMWVIudX4im4nLtkh7fQ/edit?usp=sharing)  
These results will go into your final portfolio and presentation. Organize your work as you go along to make it easier to compile later. Create slides and/or a presentation (.ppt) about your emerging data story.     
   
In case the dataset is too large to commit to GitHub, please include a link to the dataset inside the Jupyter Notebook.  
  
Discuss these results with your mentor at the next call. if you’re having trouble with your code for this unit, you can reach out to your course TA for help by emailing projects@springboard.com, or post questions in the community forum.

COMPLETE

**Unit 8.1**

**Project: Frequentist Statistics**

https://www.springboard.com/assets/images/icon-clock@2x.pngStudents typically spend 2 - 3 Hours

Ready to try your hand at frequentist inference? We’ve put together two frequentist mini-projects for you to work through. These two Jupyter Notebook-based mini-projects are contained in one project file (which you’ll find right below this text.) Please follow these steps:

1. **Download the project file and work through the mini-project called Frequentist Inference 1A.** Think of this assignment as both a demonstration and mini-project — you’ll be introduced to concepts like t- and z-tests, how to calculate standard deviation from a sample, and the Central Limit Theorem, and get some hands-on practice working with these concepts using simulated normal data. **Complete your work in your own notebook and submit it to your GitHub account.**
2. **Work through Frequentist Inference 1B.** This mini-project was designed to give you additional practice using frequentist inference with real-world data from a hospital. You’ll use this same hospital dataset while working on the other two mini-projects in this unit. **Complete your work in your own notebook and submit it to your GitHub account.**

Please review the [**rubric**](https://docs.google.com/document/d/1A2MV1HuFj-s5Gri2TRQrJIflbCaaVKrwBR33ZAZTT3k/edit) for this project.

[Download project file(s) here.](https://www.springboard.com/archeio/download/ec93bb495f29458886570593f092e1aa/)

Complete

**Unit 8.2**

**Project: Bootstrap Statistics**

https://www.springboard.com/assets/images/icon-clock@2x.pngStudents typically spend 2 - 3 Hours

For this mini-project, you’ll use the same hospital dataset you used for the frequentist inference mini-project (don’t worry — we’ve included it in the project file listed below.) You’ll analyze medical charge data to answer business questions and see firsthand how the bootstrap method can repeat the same tests you did in the frequentist mini-project (and yield consistent answers), but you’ll also use the approach to answer a question you couldn’t answer with frequentist inference.

1. Download the project file and work through the Jupyter Notebook.
2. Create your own Jupyter Notebook to log your answers and submit it to your GitHub account.

Please review the [rubric](https://docs.google.com/document/d/1adH4U0VaSsxISyYek9UISfWA_r8sN73kzFgPUlXT2lY/edit) for this project before you begin.

[Download project file(s) here.](https://www.springboard.com/archeio/download/5c75862ee63c4fcfaea7262955cba7ac/)

**Unit 8.3**

<http://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/>

**Please note that you only need to read chapters 1, 2, and 3 of this resource.**

This resource introduces you to the world of Bayesian inference and will help you get set up working with PyMC3, a useful inference tool. We recommend that you clone the associated GitHub repository and work through the first three PyMC3 chapters; this will allow you to read the notebooks like the chapters of a book, but also execute them.

There are a couple of exercises in this resource that you can also try out.   
  
**If you find that your versions of numpy become incompatible after installing PyMC3, this**[**package-list.txt file**](https://www.springboard.com/archeio/download/317b5f35c4f6435e8304325c45f02c24/) **should help you install a conda environment with compatible versions.**

**Project: Bayesian Inference**

https://www.springboard.com/assets/images/icon-clock@2x.pngStudents typically spend 2 - 3 Hours

For this final mini-project, you’ll return to the hospital dataset you’ve worked with in the previous subunits to practice applying what you’ve learned about Bayesian inference. As with the last mini-projects, we’ve included some additional tips and information to help you better understand this approach to inference.  
Again, **if you find that your versions of numpy become incompatible after installing PyMC3, this**[**package-list.txt file**](https://www.springboard.com/archeio/download/317b5f35c4f6435e8304325c45f02c24/)**should help you install a conda environment with compatible versions.**Project Steps

1. Download the project file below and work through the Jupyter Notebook.
2. Create your own Jupyter Notebook to log your answers and submit it to your GitHub account.

Please review the [rubric](https://docs.google.com/document/d/1XcsmrrM6EWXKP2L8NIHurkWI0kTnQoV00cW-FfXCmNY/edit) for this project before you begin.

[Download project file(s) here.](https://www.springboard.com/archeio/download/53a6a822459f49478f7dd417a1f46b27/)

**Unit 8.4**

<https://www.shopify.com/blog/the-complete-guide-to-ab-testing>

<https://blog.hubspot.com/marketing/a-b-testing-experiments-examples>

<https://www.ted.com/talks/esther_duflo_social_experiments_to_fight_poverty>

**Unit 8.5**

**Project: Capstone Project 1: Statistical Data Analysis**

https://www.springboard.com/assets/images/icon-clock@2x.pngStudents typically spend 4 - 12 Hours

At this point, you’ve obtained the dataset for your capstone project, cleaned, and wrangled it into a form that's ready for analysis. It's now time to apply the inferential statistics techniques you’ve learned to explore the data.

Based on your dataset, the questions that interest you, and the results of the visualization techniques that you used previously, you should choose the most relevant statistical inference techniques. You aren’t expected to demonstrate all of them. Your specific situation determines how much time it’ll take you to complete this project. Talk to your mentor to determine the most appropriate approach to take for your project. You may find yourself revisiting the analytical framework that you first used to develop your proposal questions. It’s fine to refine your questions more as you get deeper into your data and find interesting patterns and answers. Remember to stay in touch with your mentor to remain focused on the scope of your project

Think of the following questions and apply them to your dataset:

* Are there variables that are particularly significant in terms of explaining the answer to your project question?
* Are there significant differences between subgroups in your data that may be relevant to your project aim?
* Are there strong correlations between pairs of independent variables or between an independent and a dependent variable?
* What are the most appropriate tests to use to analyze these relationships?

**Submission**: Write a 1-2 page report on the steps and findings of your inferential statistical analysis. Upload this report to your GitHub and submit a link. Eventually, this report will get incorporated into your milestone report.

You will be evaluated using this [rubric](https://docs.google.com/document/d/1sRNBAj1zf7uLAJLPBTFAxzsH3Qe31kAwCSEvazIz-uE/edit?usp=sharing).

**Unit 8.6**

**Khan Academy Videos Optional**

The following resources were all created by Khan Academy and explore a variety of topics, some of which have already been touched on in this unit.

1. [Random Variables](https://www.khanacademy.org/math/statistics-probability/random-variables-stats-library) —  This video teaches you about random variables, which can be any outcomes from a chance process, such as the number of times 'heads' appears when flipping a coin 20 times. When working on a statistics problem, it's good to know how to calculate probabilities and expected values for random variables.

**Done**

1. [Sampling Distributions](https://www.khanacademy.org/math/statistics-probability/sampling-distributions-library) —  This exercise on a sampling distribution shows the frequency and result a statistic can take in any sample from a population. This topic covers how proportions and means behave in repeated samples.

**Done**

1. [One Sample Confidence Interval](https://www.khanacademy.org/math/statistics-probability/confidence-intervals-one-sample)— This course covers confidence intervals for means and proportions.

**Done**

1. [One Sample Significance Test](https://www.khanacademy.org/math/statistics-probability/significance-tests-one-sample) — This course explores significance tests, which provide a formal process for using sample data to evaluate the likelihood of some claim about a population value. You'll learn how to calculate p-values to see how likely a sample result is to occur by chance and to make conclusions about hypotheses.
2. [Two-Sample Inference for the Difference Between Groups](https://www.khanacademy.org/math/statistics-probability/significance-tests-confidence-intervals-two-samples)— In this course, you’ll learn how to apply your knowledge of confidence intervals and significance tests to situations that involve comparing two samples to see if there is a significant difference between two populations.
3. [Inference for Categorical Data](https://www.khanacademy.org/math/statistics-probability/inference-categorical-data-chi-square-tests) — Chi-square tests are a family of significance tests that provide ways to test hypotheses about distributions of categorical data. This topic covers goodness-of-fit tests to see if sample data fits a hypothesized distribution and tests for independence between two categorical variables.

Interview Questions:

<https://docs.google.com/document/d/1cHBwjt0bWEoO5tcaYx5c5t_GOYr4fbbT_k-bCLgSoGo/edit?usp=sharing>

**Unit 9**

Informational Interview

**Unit 10.1**

**10.11 Bias and Regression**

DONE

**10.12 Regression Cont.**

DONE

**10.13 Classification**

DONE

<https://www.datacamp.com/courses/supervised-learning-with-scikit-learn>

DONE

### Project: Linear Regression Using Boston Housing Data Set

https://www.springboard.com/assets/images/icon-clock@2x.pngStudents typically spend 2 - 4 Hours

**Instructions:**

Please download and open the zipped file and work in the Jupyter Notebook in the unzipped directory. In the notebook, the phrase “Your turn” indicates sections where you need to fill in the code. After you’re done, please add the entire directory on your GitHub and submit a link to the completed Jupyter Notebook.

[Download the Mini-Project Evaluation Rubric here](https://drive.google.com/open?id=1bUiXAxqyprjSkNwx2GfWpnqo4jHPSZ1NCwsNCg-RDtg)

[Download project file(s) here.](https://www.springboard.com/archeio/download/18350417e3834bbd8ed17f66fec496c3/)

### Project: Heights and Weights Using Logistic Regression

https://www.springboard.com/assets/images/icon-clock@2x.pngStudents typically spend 2 - 4 Hours

**Instructions:**  
  
Please download and open the zipped file and work in the Jupyter Notebook in the unzipped directory. In the notebook, the phrase “Your turn” indicates sections where you need to fill in the code. After you're done, please add the entire directory on your GitHub and submit a link to the completed Jupyter Notebook.

[Download the Mini-Project Evaluation Rubric here](https://drive.google.com/open?id=1QP5hnEEQKaOm6IJ2UnIpHr5gJj3X0SnEejBBsCoATLk)

[Download project file(s) here.](https://www.springboard.com/archeio/download/e15045f38bb54344a952bef3ed291540/)

**Unit 10.2**

SVM and Trees

<https://matterhorn.dce.harvard.edu/engage/player/watch.html?id=f21fcc8f-93a8-49f6-9ff8-0f339b0728bd>

**DONE**

Decision Trees

<https://matterhorn.dce.harvard.edu/engage/player/watch.html?id=8892a8b7-25eb-4bc5-80b6-47b9cf681a05>

DONE

Using Random Forests in Python

<https://www.youtube.com/watch?v=6O4kASc-SDE&feature=youtu.be>

Ensemble Methods

<https://matterhorn.dce.harvard.edu/engage/player/watch.html?id=4831ebf0-7832-42c5-9339-5b5e08dd3e92>

<https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d>

**Unit 10.3**

10.31 Bayesian

DONE

Video 2

<https://youtu.be/y3ZTKFZ-1QQ>

### Project: Predicting Movie Ratings from Reviews Using Naive Bayes

https://www.springboard.com/assets/images/icon-clock@2x.pngStudents typically spend 2 - 5 Hours

**Instructions:**  
  
Please download and open the zipped file and work in the Jupyter Notebook in the unzipped directory. In the notebook, the phrase “Your Turn" indicates sections where you need to fill in the code. After you're done, please add the entire directory on your github and submit a link to the completed Jupyter Notebook.  
 [Download the Mini-Project Evaluation Rubric here](https://drive.google.com/open?id=1YIouTCz0Jlg2wJUOUCybT9QMonfEPgg3C6z2_wwj36E)

[Download project file(s) here.](https://www.springboard.com/archeio/download/02471e657fa246fa8a2b28ad45fdd21d/)

**Unit 10.4**

[**https://www.datacamp.com/courses/machine-learning-with-the-experts-school-budgets**](https://www.datacamp.com/courses/machine-learning-with-the-experts-school-budgets)

**DONE**

Stopping here for material